Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry 14 (2006) 6933–6939
Bioorganic & Medicinal Chemistry

QSAR modeling of anti-invasive activity of organic compounds using structural descriptors

Alan R. Katritzky,[a,*] Minati Kuanar,[a] Dimitar A. Dobchev,[a,b] Barbara W. A. Vanhoecke,[c,*] Mati Karelson,[b] Virinder S. Parmar,[d] Christian V. Stevens[e] and Marc E. Bracke[f,*]

[a]*Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, FL 32611, USA*
[b]*Department of Chemistry, University of Tartu, 2 Jakobi street, Tartu 51014, Estonia*
[c]*Department of Gynaecological Oncology, University Hospital, Ghent University, De Pintelaan 185, B-9000 Gent, Belgium*
[d]*Bioorganic Laboratory, Department of Chemistry, University of Delhi, Delhi 110007, India*
[e]*Department of Organic Chemistry, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, B-9000 Gent, Belgium*
[f]*Laboratory of Experimental Cancerology, Department of Radiotherapy and Nuclear Medicine, University Hospital, Ghent University, De Pintelaan 185, B-9000 Gent, Belgium*

Received 16 March 2006; revised 14 June 2006; accepted 19 June 2006

Abstract—The anti-invasive activity of 139 compounds was correlated by an artificial neural network approach with descriptors calculated solely from the molecular structures using CODESSA Pro. The best multilinear regression method implemented in CODESSA Pro was used for a pre-selection of descriptors. The resulting nonlinear (artificial neural network) QSAR model predicted the exact class for 66 (71%) of the training set of 93 compounds and 32 (70%) of validation set of 46 compounds. The standard deviation ratios for the both training and validation sets are less than unity, indicating a satisfactory predictive capability for classification of the nature of the anti-invasive activity data. The proposed model can be used for the prediction of the anti-invasive activity of novel classes of compounds enabling a virtual screening of large databases of anticancer drugs.

1. Introduction

'Invasion' is a measure of tumor cell activity, closely related to the final outcome of cancer. The development of anti-invasive and anti-metastasis drugs is a major challenge in current cancer research. A number of natural products have been found to exhibit cytotoxic activity against tumor cell lines. Among them a few flavonoids inhibit tumor invasion in vitro (3,7-dimethoxyflavone[1] and (+)-catechin)[2] and artificial metastasis in vivo[3] (flavone acetic acid). Anti-invasive activity of selected alkaloids and polyphenolics,[4] 1,3-diarylpropenones (chalcones)[5] has been studied and found to inhibit growth and invasion of human mammary carcinoma cells with fragments of embryonic chick heart in vitro. The Bracke research group[6] tested the activity of 95 compounds of diversified classes against invasive human (MCF-7/6) mammary carcinoma cells confronting embryonic chick heart fragments in vitro. Among polyphenolic and heterocyclic compounds, the authors observed that a pyrazole derivative, an isoxazolylcoumarin, and a prenylated desoxybenzoin inhibited invasion at concentrations of 1 μM without being cytotoxic for the cancer cells.[6] The same group also studied the inhibition of invasion by selective alkaloids and polyphenolic compounds such as tangeretin, xanthohumol, prenylated chalcones, prenylated desoxybenzoin pyrazoles, isoxazolylcoumarin, and a prenylated desoxybenzoin against human MCF-7/6 mammary carcinoma cells with embryonic chick heart fragments.[7] The tea polyphenols, epicatechin gallate (ECG), epigallocatechin gallate (EGCG), and theaflavin, flavonoids, and some antioxidants [8–11] have shown their potential inhibitions against tumor cells. Some synthetic coumarin derivatives and hydrolysable tannins were found to inhibit tumor metastasis.[12,13] The structural factors which relate the anticancer activity of these compounds are still not clear.

Keywords: Anti-invasive activity; QSAR; Molecular descriptors; Artificial neural network; Polyphenols.
* Corresponding authors. Tel.: +1 352 392 554; fax: +1 352 392 9199 (A.R.K); e-mail: katritzky@chem.ufl.edu

0968-0896/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2006.06.036

Recently, quantitative structure–activity relationships (QSARs) have been used extensively to develop models in order to estimate and predict biological or toxicological behavior of organic molecules using computational descriptors solely derived from chemical structures. The use of computational tools for prediction of the activity of analogs of drug candidates prior to their synthesis has accelerated the drug discovery processes in the pharmaceutical industry. Computational approaches have been successful in predicting various ADME properties including blood–brain barrier penetration or intestinal absorption,[14–17] and binding affinity data.[18,19] The search for anticancer compounds has always been a priority of medicinal chemists; computer-aided drug design approaches (CADD)[20] have emerged as a promising tool in the search for new lead compounds.[21] Recently, Jorgensen reviewed the many roles of computation in drug discovery.[22]

Estrada's research group[23] has applied a novel approach for the selection and design of anticancer compounds. The authors developed a QSAR model by using a topological substructural approach to molecular design (TOSS-MODE) which discriminates anticancer compounds from the inactive ones in a training series.[23] Neural networks have been widely used for classification and for approximation in various fields of chemistry and bioinformatics.[24] Xue et al.[25] have successfully used the probabilistic neural networks for the classification of 102 active compounds from diverse medicinal plants with anticancer activity using molecular descriptors. The PNNs results are found to be superior than the linear discriminant analysis.[25]

Our main aim in the present study is to develop a structure–activity relationship that could help to predict novel classes of compounds having potential anti-invasive activity.

2. Data set

The anti-invasive activity data for 139 compounds were taken from Ref. 7. Since no continuous activity values were measured in this test, the compounds possessing activity at 1, 10, and 100 μM concentrations were classified into the following categories (cf. Table 1).

These anti-invasive activity score (I_index) data were taken as dependent variables for the correlation study and are given in Supplementary information (see SM-1).

2.1. Methodology

2.1.1. Molecular modeling. The structures of the compounds were drawn using ISIS/Draw as implemented in the ISIS 2.4 package and pre-optimized using molecular mechanics force fields (MM+) encoded in HyperChem software.[26] The molecular geometries were refined using AM1 Hamiltonian (Austin Method 1)[27] calculations together with eigenvector following a geometry optimization procedure available in the quantum chemical program MOPAC 7.05, implemented in the CODESSA Pro package.[28] The gradient norm criterion 0.01 kcal/Å was applied in the geometry optimization for all structures. These optimized structures were loaded in CODESSA Pro and more than 800 theoretical descriptors were calculated. These descriptors can be classified into several groups: (i) constitutional, (ii) topological, (iii) geometrical, (iv) thermodynamic, (v) quantum chemical, and (vi) charge-related.

Table 1. Classification of the anti-invasive activity data

Concentration (μM)	Activity	Anti-invasive activity score (I_index)
>100	Low	1
100	Fair	2
10	Good	3
1	Active	4

2.1.2. Linear approach. CODESSA Pro[28] includes diverse statistical structure–property–activity correlation techniques that can be used for the analysis in combination with the calculated molecular descriptors. CODESSA Pro enables the calculation of numerous descriptors solely on the basis of molecular structural information.[29–31] Since only theoretically calculated descriptors are used in the resulting multiparameter correlation equations, the value of the property/activity of interest can be predicted for an unknown structure. CODESSA Pro methodology has given promising results in successive prediction of HIV-1 protease inhibitory activity of substituted tetrahydropyrimidinones,[32] antibacterial activity of 3-aryloxazolidine-2-one,[33] the binding energies for 1:1 complexation systems between various organic guest molecules and β-cyclodextrin,[34] blood and tissue air partition coefficients of organic solutes,[35,36] prediction of partition of drugs in human milk and plasma,[37] antimalarial activity of drugs,[38] inhibition of the platelet-derived growth factor of 1-phenylbenzimidazoles,[39] and blood–brain distribution coefficients of drugs.[40]

2.1.3. Linear modeling. An important stage of the multilinear regression QSAR methodology is the search for the best multilinear equation among a given pool of descriptors. Eq. 1 gives the mathematical representation of the equation that should correlate the best inhibitory activity (A) with a certain number n of molecular descriptors (D_i) weighted by the regression coefficients b_i:

$$A = b_0 + \sum_{i=1}^{n} b_i D_i \quad (1)$$

The best multilinear regression method (BMLR)[41,42] encoded in CODESSA Pro software was used to select significant descriptors for building multilinear QSAR models. The treatment started with the reduction of the number of molecular descriptors. If two descriptors were highly correlated, then only one descriptor was selected; the descriptors with insignificant variance were also rejected. This helps to speed up the descriptor selection and reduces the probability of including unrelated descriptors by chance.

The strategy used to develop physically meaningful multilinear QSAR equations from the very large pool

of descriptors is a combination of the multilinear regression and forward selection procedures. This strategy involved the following steps:

(1) Detection of all orthogonal pairs of descriptors i and j from the given descriptor space. Pairs of descriptors with a correlation coefficient $R^2_{ij} > 0.5$ were considered intercorrelated and such pairs were eliminated at this stage.
(2) From the complete set of all the two-parameter regression equations of orthogonal pairs the 400 possessing the highest R^2 value two-parameter equations were used.
(3) Search for superior multiparameter regression equations: for each descriptor pair, retained in the previous step, additional noncollinear descriptor vectors were successively added, and the appropriate $(n + 1)$-parameter regression treatment was carried out. When the Fisher criterion F (or cross-validation coefficient R_{cv}) obtained for any of these correlations was lower than for the best correlation of the previous rank (n), the latter was designed as the final result and the search was terminated. Otherwise, the descriptor sets with the highest coefficient of determinations were stored and the current step was repeated with the number of parameters (descriptors) increased by one $(n + 2)$.

The final result had therefore the maximum value of the Fisher criterion and the highest cross-validated coefficient of determination.

A major decision in developing successive QSAR is when to stop adding descriptors to the model during the stepwise regression procedure. A simple technique to control the model expansion is the so-called 'breaking point' in the improvement of the statistical quality of the model, by analyzing the plot of the number of descriptors involved in the obtained models versus squared-correlation coefficient values corresponding to those models. Frequently, the statistical improvement of the regression model is less significant ($\Delta R^2 < 0.02$) after a certain number of independent variables in the model ('breaking point'). Consequently, the model corresponding to the breaking point is considered the best/optimum model.

2.1.4. Nonlinear ANN approach. Artificial neural networks (ANNs)[43–45] have become an important modeling technique in numerous areas of chemistry and pharmacy.[46–48] The mathematical adaptability of ANN commends them as a powerful tool for pattern classification and building predictive models. A particular advantage of ANNs is their inherent ability to incorporate nonlinear dependencies between the dependent and independent variables without using an explicit mathematical function.

The present study involves two different approaches to correlate the anti-invasive activity score data for a large number of drug-like molecules with the structural descriptors; (i) QSAR modeling, by multilinear regression performed with the CODESSA Pro program which applies up to 863 different constitutional, geometrical, topological, electrostatic, quantum chemical, and thermodynamic molecular descriptors, and (ii) nonlinear modeling, performed using artificial neural networks (ANN) with back propagation learning algorithm and sigmoid activation function developed in-house. In this work, a backpropagation network[49,50] was developed and used to obtain a nonlinear QSAR model. Topologically, it consists of input, hidden, and output layers of neurons or units connected by weights. Each input layer node corresponds to a single independent variable (molecular descriptor) with the exception of the bias node. Similarly, each output layer node corresponds to a different dependent variable (property under investigation). In both of these treatments all descriptors used are derived solely from molecular structure and do not require experimental data or expensive theoretical calculations to be obtained.

Here, we show that the combination of these two different approaches (multilinear and nonlinear) led to pertinent QSAR models, and their joint application improves the robustness of predictions.

3. Results and discussions

3.1. Linear modeling for anti-invasive activity score (I_index)

As the anti-invasive activity score (I_index) data possess reduced number of discrete values, the multilinear techniques were found not appropriate for modeling. However, we used the BMLR method implemented in CODESSA Pro to select a set of descriptors which are used as inputs in the ANN model. The descriptor set selected is the best linear combination in the descriptor space consisting of 863 variables. Then nonlinear transformation (ANN) was applied to improve the significance of the predictive QSAR model. The best seven descriptors in the BMLR equation with $R^2 = 0.374$ were used as input parameters for ANN modeling in the next section.

3.2. Nonlinear modeling for anti-invasive activity score (I_index)

In this study, we used backpropagation ANN methodology for classification of the anti-invasive activity score (I_index). This index has four classes (discrete values) that were used as a property under investigation for our nonlinear ANN modeling.

The experimental values (139 data points) were divided into two sets: training and validation sets. The training and validation sets consist of 93 and 46 data points, respectively. These two sets were randomly reordered and selected. The reason for this division is to avoid overfitting[45] of the neural network model since it has the capability to learn even the experimental noise of the data.

To build a reliable neural network model, a selection of the input descriptors was performed based on the

BMLR algorithm encoded in CODESSA Pro. Since, this algorithm possesses rigorous criteria (squared intercorrelation coefficient of the descriptors $R^2_{int} < 0.6$, variance of the descriptors var $< 10^{-3}$, R^2 of the equation, cross-validated squared correlation coefficient R^2_{cv}, randomization test, etc.) for descriptor selection used in multilinear model, we selected the best seven descriptors of multilinear equation obtained by BMLR as inputs for the NN model. However, since the BMLR searches only for linear relationship, we used also nonlinear combination of these independent variables (descriptors). These were selected by exploring the scatter plots with the property. Some of the descriptors (18) showing extended nonlinear dependence with the invasive index were selected for further treatment. Thus, several nonlinear functions were automatically constructed from these descriptors as square, square root and binomes. By doing this we extended the whole descriptor space including nonlinearities which could lead to better ANN models. This space was reduced according to the criterion for small variance ratio of the descriptors, that is $\sigma/|d_{max} - d_{min}| < 10^{-3}$ were excluded (311). Further, the BMLR was used to select the appropriate descriptors as inputs from the reduced space. The selected seven descriptors resulted in the above method were: HA-dependent HDCA-2 (Zefirov PC), *d*1; WNSA-3 Weighted PNSA (PNSA3*TMSA/1000) (MOPAC PC), *d*2; (minimum atomic orbital electronic population) × (minimum e–e repulsion for atom H), *d*3; maximum resonance energy for bond C–C, *d*4; Max e–n attraction for bond C–C, *d*5; maximum Coulombic interaction for bond H–C, *d*6; and minimum e–e repulsion for atom H, *d*7. The descriptor *d*3 is constructed by multiplication of two descriptors that was significant contribution to the nonlinear dependence with the property. The BMLR equation chosen for the descriptor selection is shown in (2):

$$\begin{aligned} I\_\text{index} = {} & 5.47 \pm 4.40 + (6.75 \pm 1.13)d6 \\ & - (0.15 \pm 0.02)d4 + (0.2 \pm 0.004)d2 \\ & + (1.35 \pm 0.35)d7 + (0.96 \pm 0.21)d1 \\ & + (5.24 \pm 1.88)d3 + (0.22 \pm 0.08)d5 \quad (2) \end{aligned}$$

$R^2 = 0.374$, $F = 11.202$, $s^2 = 0.481$.

For the training of the NN model the backpropagation algorithm with Levenberg Marquardt optimizer[49] was

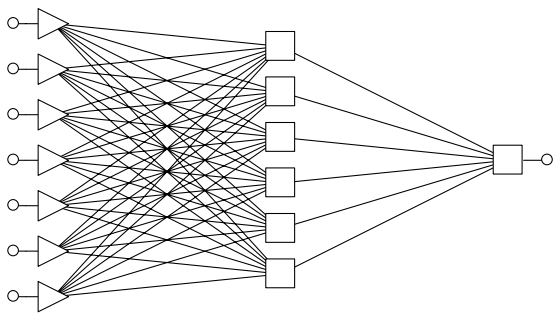

Figure 1. The best ANN architecture for the model (7–6–1).

Table 2. Experimental and predicted classes for the training and validation sets according to the ANN model

Compound	Training set	
	Predicted class	Experimental class
3	1	1
4	1	1
5	2	1
6	1	1
7	4	4
8	3	3
9	1	1
11	2	1
12	1	3
15	3	3
21	2	1
22	3	3
23	1	1
25	1	2
26	2	2
27	3	4
28	3	3
29	1	1
30	4	4
32	2	1
34	1	1
35	2	1
36	2	2
37	4	4
38	2	1
39	1	1
40	3	3
41	4	4
52	3	4
53	4	4
54	4	4
56	4	4
58	3	3
59	3	3
60	2	3
61	3	3
62	3	3
63	3	3
64	3	3
65	3	3
67	2	3
70	3	3
72	3	3
73	3	3
74	3	3
75	3	3
77	3	3
79	3	3
80	3	3
82	2	3
84	3	3
86	3	3
87	2	3
90	3	3
91	3	3
92	2	3
93	3	3
94	3	3
96	3	3
97	3	3
98	2	2
99	2	2
100	3	2
102	1	2

Table 2 (*continued*)

Compound	Training set	
	Predicted class	Experimental class
104	2	2
106	2	2
108	2	2
109	3	2
110	2	2
112	3	2
113	3	2
116	3	2
117	3	2
118	2	2
119	3	2
120	2	2
121	3	2
122	2	2
125	2	2
126	3	2
127	3	2
128	2	2
129	2	2
131	2	2
132	2	2
135	2	2
136	2	2
138	3	2
142	2	2
146	2	2
147	2	2
148	2	2
150	2	2
Compound	Validation set	
	Predicted class	Experimental class
1	1	1
2	1	1
10	1	1
13	2	1
14	1	2
16	1	1
17	1	1
18	1	1
19	3	3
20	1	1
24	2	1
31	1	1
33	2	2
48	3	3
55	4	4
57	4	3
66	3	3
68	3	3
69	3	3
71	3	3
76	3	3
78	4	4
81	3	3
83	2	3
85	2	3
88	2	3
89	3	3
95	3	3
101	1	2
103	2	2
105	3	2
107	3	2

Table 2 (*continued*)

Compound	Validation set	
	Predicted class	Experimental class
111	3	2
114	3	2
115	3	2
123	2	2
124	2	2
130	2	2
133	2	2
134	2	2
137	2	2
139	2	2
140	3	2
141	2	2
145	2	2
149	2	2

used. The ANN model was trained only on the training set since the validation set was used to monitor the external prediction error (rms) and thus to avoid overtraining. Among the 11 architectures constructed the best ANN architecture we found was 7–6–1 as shown in Figure 1. That is, in the first layer seven inputs comprised of seven input descriptors, hidden layer comprised of six neurons, and the last output layer comprised of the one neuron for the property modeled.

After the training of the ANN, the optimized weights were set in the network. The prediction results based on this ANN model are shown in Table 2. A graphical presentation between the experimental and predicted data according to the ANN model is given by their confusion matrices shown in Figures 2 and 3 for the training and validation sets, respectively.

The statistical criteria obtained for the ANN model for both training and validation sets are shown in Table 3. As can be seen from this table the rms error for the training set is quite low. In addition, the rms for the validation set is also low showing the good prediction ability. Data mean criterion for both sets is very close to each other, that is, the overall prediction is close to experimental. Also, taking into account that the SD ratio (showing stability of the model) is less than one for both sets we can conclude that the ANN model is satisfactorily predictive having in mind the classification nature of the experimental data.

As can be seen from Figure 2, 57% of the class 1 was predicted correctly. For the remaining classes, the prediction is as follows: class 2—63%, class 3—82%, and class 4—78%. Consequently, the average percentage to predict the exact class is 71% for the training set, which is quite significant. With respect to the validation set (which serves as an external set, see Fig. 3) the exactly predicted probabilities are: class 1—80%, class 2—60%, class 3—71%, and class 4—100%. From both figures it is noticeable that the largest number of compounds exactly predicted is situated on the main left diagonal of the confusion matrices. It can be noted that class 4 for the validation set is predicted exact-

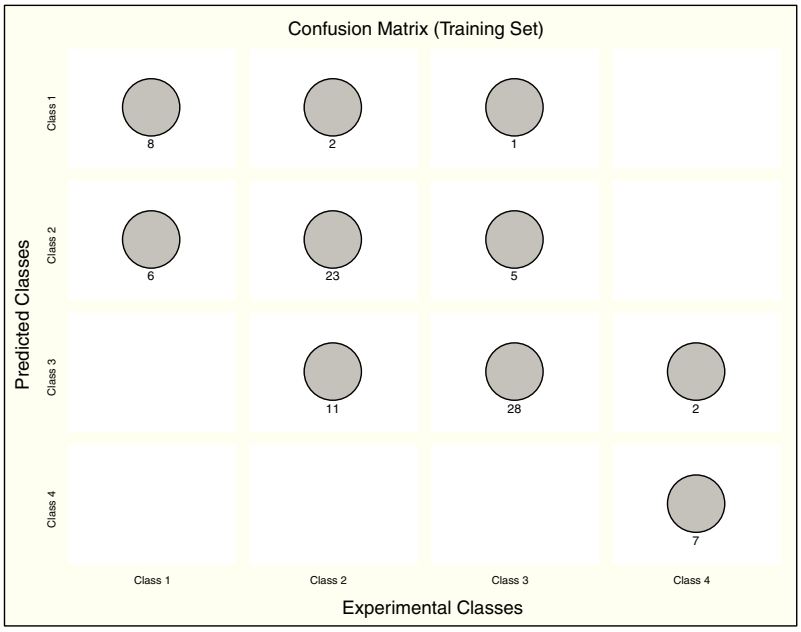

Figure 2. Confusion matrix for the training set (93 points) according to the ANN model.

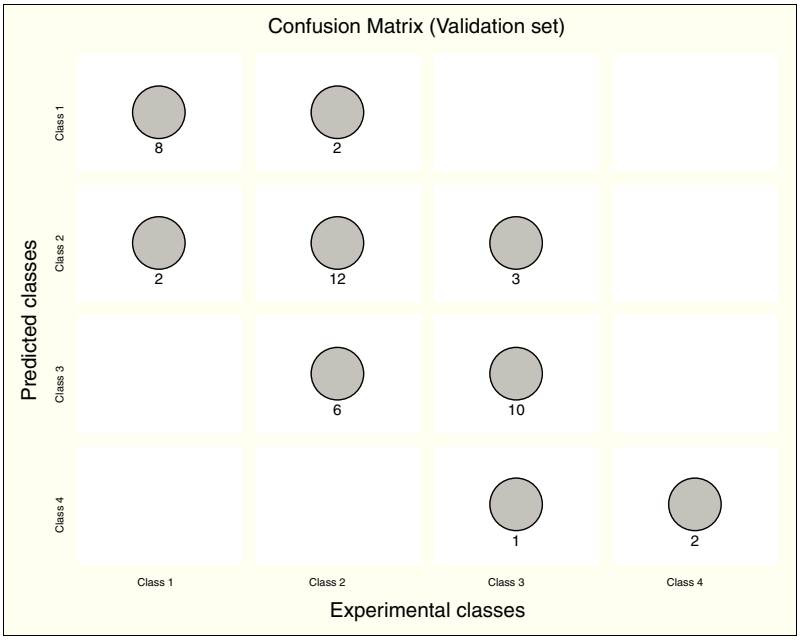

Figure 3. Confusion matrix for the validation set (46 points) according to the ANN model.

ly that is goal of the most practical cases where one needs to know which compound is most active among the novel compounds synthesized.

Table 3. Statistical characteristics of ANN model

Criterion	Training (exp/pred)	Validation (exp/pred)
Data mean	2.409/2.473	2.174/2.261
Data SD	0.863/0.802	0.825/0.880
Abs error mean	0.301	0.304
Error SD	0.484	0.465
rms	0.568	0.569
SD ratio	0.561	0.564
Correlation	0.770	0.793

4. Conclusions

The anti-invasive activity score (I_index) data were correlated with theoretically calculated molecular descriptors through an appropriately trained artificial neural network. A multilinear QSAR model was initially developed for the purposes of descriptor selection. Notably, all descriptors appearing in the seven-parameter regression equation in the ANN model have been derived from theoretical descriptors calculated from chemical structure. The current computational power available for chemical research allows such calculations for large data sets in realistic time. The descriptors used in the model relate to the essential electrostatic, conformation-

al interactions, and hydrogen acceptor/donor abilities in the biological system. Thus, the ANN model developed in the present work can be used for the prediction of anti-invasive activity of novel class of compounds.

Acknowledgments

The authors thank the reviewers for helpful suggestions and comments on the manuscript.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2006.06.036.

References and notes

1. Parmar, V. S.; Jain, R.; Sharma, S. K.; Vardhan, A.; Jha, A.; Taneja, P.; Singh, S.; Vyncke, B. M.; Bracke, M. E.; Mareel, M. M. *J. Pharm. Sci.* **1994**, *83*, 1217.
2. Bracke, M. E.; Van Cauwenberge, R. M. L.; Mareel, M. M. *Clin Exp. Metastasis* **1984**, *2*, 161.
3. Giavazzi, R.; Garofalo, A.; Damia, G.; Garattini, S.; D'Incalci, M. *Br. J. Cancer* **1988**, *57*, 277.
4. Parmar, V. S.; Bracke, M. E.; Philippe, J.; Wengel, J.; Jain, S. C.; Olsen, C. E.; Bisht, K. S.; Sharma, N. K.; Courtens, A.; Sharma, S. K.; Vennekens, K.; Van Marck, V.; Singh, S. K.; Kumar, N.; Kumar, A.; Malhotra, S.; Kumar, R.; Rajwanshi, V. K.; Jain, R.; Mareel, M. M. *Bioorg. Med. Chem.* **1997**, *5*, 1609.
5. Mukherjee, S.; Kumar, V.; Prasad, A. K.; Raj, H. G.; Bracke, M. E.; Olsen, C. E.; Jain, S. C.; Parmar, V. S. *Bioorg. Med. Chem.* **2001**, *9*, 337.
6. Parmar, V. S.; Sharma, N. K.; Husain, M.; Watterson, A. C.; Kumar, J.; Samuelson, L. A.; Cholli, A. L.; Prasad, A. K.; Kumar, A.; Malhotra, S.; Kumar, N.; Jha, A.; Singh, A.; Singh, I.; Himanshu; Vats, A.; Shakil, N. A.; Trikha, S.; Mukherjee, S.; Sharma, S. K.; Singh, S. K.; Kumar, A.; Jha, H. N.; Olsen, C. E.; Stove, C. P.; Bracke, M. E.; Mareel, M. M. *Bioorg. Med. Chem.* **2003**, *11*, 913.
7. Vanhoecke, B. W.; Depypere, H. T.; De Beyter, A.; Sharma, S. K.; Parmar, V. S.; De Keukeleire, D.; Bracke, M. E. *Pure Appl. Chem.* **2005**, *77*, 65.
8. Maeda-Yamamoto, M.; Kawahara, H.; Tahara, N.; Tsuji, K.; Hara, Y.; Isemura, M. *J. Agric. Food Chem.* **1999**, *47*, 2350.
9. Isemura, M.; Saeki, K.; Kimura, T.; Hayakawa, S.; Minami, T.; Sazuka, M. *Biofactors* **2000**, *13*, 81.
10. Sartor, L.; Pezzato, E.; Dell'Aica, I.; Caniato, R.; Biggin, S.; Garbisa, S. *Biochem. Pharmacol.* **2002**, *64*, 229.
11. Hou, Z.; Lambert, J. D.; Chin, K-V.; Yang, C. S. *Mutat. Res.* **2004**, *555*, 3.
12. Kempen, I.; Papapostolou, D.; Thierry, N.; Pochet, L.; Counerotte, S.; Masereel, B.; Foidart, J.-M.; Reboud-Ravaux, M.; Noel, A.; Pirotte, B. *Br. J. Cancer* **2003**, *88*, 1111.
13. Tanimura, S.; Kadomoto, R.; Tanaka, T.; Zhang, Y-J.; Kouno, I.; Kohno, M. *Biochem. Biophys. Res. Commun.* **2005**, *330*, 1306.
14. Clark, D. E. *Comb. Chem. High Throughput Screening* **2001**, *4*, 477.
15. Atkinson, F.; Cole, S.; Green, C.; van de Waterbeemd, H. *Curr. Med. Chem.* **2002**, *2*, 229.
16. Norinder, U.; Haeberlein, M. *Adv. Drug Delivery Rev.* **2002**, *54*, 291.
17. Clark, D. E. *Drug Discovery Today* **2003**, *8*, 927.
18. Beteringhe, A.; Filip, P.; Tarko, L. *ARKIVOC* **2005**, *x*, 45.
19. Ramamurthi, N.; Gunturi, S. B. *ARKIVOC* **2004**, *xi*, 102.
20. Buchwald, P.; Bodor, N. *Drugs Future* **2002**, *27*, 577.
21. Lunney, E. A. *Med. Chem. Res.* **1998**, *8*, 352.
22. Jorgensen, W. L. *Science* **2004**, *303*, 1813.
23. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. *J. Med. Chem.* **2000**, *43*, 1975.
24. (a) Schneider, G.; Wrede, P. *Prog. Biophys. Mol. Biol.* **1998**, *70*, 175; (b) Lucic, B.; Trinajstic, N. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 121; (c) Lucic, B.; Nadramija, D.; Basic, I.; Trinajstic, N. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1094.
25. Xue, C. X.; Zhang, X. Y.; Liu, M. C.; Hu, Z. D.; Fan, B. T. *J. Pharm. Biomed. Anal.* **2005**, *38*, 497.
26. *Hyperchem*, v. 7.5. Hypercube Inc.; Gainesville, FL.
27. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
28. CODESSA Pro Software, University of Florida, 2002.
29. Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *Chem. Soc. Rev.* **1995**, *24*, 279.
30. Katritzky, A. R.; Karelson, M.; Lobanov, V. *Pure Appl. Chem.* **1997**, *69*, 245.
31. Katritzky, A. R.; Fara, D. C.; Petrukhin, R. O.; Tatham, D. B.; Maran, U.; Lomaka, A.; Karelson, M. *Curr. Top. Med. Chem.* **2002**, *2*, 1333.
32. Katritzky, A. R.; Oliferenko, A.; Lomaka, A.; Karelson, M. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 3453.
33. Katritzky, A. R.; Fara, D. C.; Karelson, M. *Bioorg. Med. Chem.* **2004**, *12*, 3027.
34. Katritzky, A. R.; Fara, D. C.; Yang, H.; Karelson, M.; Suzuki, T.; Solov'ev, V. P.; Varnek, A. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 529.
35. Katritzky, A. R.; Kuanar, M.; Fara, D. C.; Karelson, M.; Acree, W. E., Jr. *Bioorg. Med. Chem.* **2004**, *12*, 4735.
36. Katritzky, A. R.; Kuanar, M.; Fara, D. C.; Karelson, M.; Acree, W. E.; Solov'ev, V. P.; Varnek, A. *Bioorg. Med. Chem.* **2005**, *13*, 6450.
37. Katritzky, A.; Dobchev, D.; Hur, E.; Fara, D.; Karelson, M. *Bioorg. Med. Chem.* **2005**, *13*, 1623.
38. Katritzky, A. R.; Kulshyn, O. V.; Stoyanova-Slavova, I.; Dobchev, D. A.; Kuanar, M.; Fara, D. C.; Karelson, M. *Bioorg. Med. Chem.* **2006**, *14*, 2333.
39. Katritzky, A. R.; Dobchev, D. A.; Fara, D. C.; Karelson, M. *Bioorg. Med. Chem.* **2005**, *13*, 6598.
40. Katritzky, A. R.; Kuanar, M.; Slavov, S.; Fara, D. C.; Dobchev, D. A.; Karelson, M.; Acree, W. E., Jr. *Bioorg. Med. Chem.* **2006**, *14*, 4888.
41. Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. *J. Phys. Chem.* **1996**, *100*, 10400.
42. Karelson, M. *Molecular Descriptors in QSAR/QSPR*; Wiley-Interscience: New York, 2000.
43. Tetteh, J.; Suzuki, T.; Metcalfe, E.; Howells, S. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 491.
44. Goll, E. S.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1081.
45. Zupan, J.; Gasteiger, J. *Neural Networks for Chemists: an Introduction*; VCH-Verlag: Weinheim, 1993.
46. Burns, J. A.; Whitesides, G. M. *Chem. Rev.* **1993**, *93*, 2583.
47. Svozil, D.; Kvasnicka, V.; Pospichal, J. *Chemom. Intell. Lab. Syst.* **1997**, *39*, 43.
48. Agatonovic-Kustrin, S.; Ling, L. H.; Tham, S. Y.; Alany, R. G. *J. Pharm. Biomed. Anal.* **2002**, *29*, 103.
49. Haykin, S. *Neural Networks. A comprehensive foundation*; Pearson, 1999.
50. Masters, T. *Practical Neural Network Recipes in C++*; Academic Press, 1993.